

STROKES LOCALIZATION BY m-ARRAY DECODING AND FAST IMAGE MATCHING

Technical Field

- [01] The present invention relates to interacting with a medium using a digital pen. More particularly, the present invention relates to determining the location of a digital pen during interaction with one or more surfaces.

Background

- [02] Computer users are accustomed to using a mouse and keyboard as a way of interacting with a personal computer. While personal computers provide a number of advantages over written documents, most users continue to perform certain functions using printed paper. Some of these functions include reading and annotating written documents. In the case of annotations, the printed document assumes a greater significance because of the annotations placed on it by the user. One of the difficulties, however, with having a printed document with annotations is the later need to have the annotations entered back into the electronic form of the document. This requires the original user or another user to wade through the annotations and enter them into a personal computer. In some cases, a user will scan in the annotations and the original text, thereby creating a new document. These multiple steps make the interaction between the printed document and the electronic version of the document difficult to handle on a repeated basis. Further, scanned-in images are frequently non-modifiable. There may be no way to separate the annotations from the original text. This makes using the annotations difficult. Accordingly, an improved way of handling annotations is needed.
- [03] One technique of capturing handwritten information is by using a pen whose location may be determined during writing. One pen that provides this capability is the Anoto pen by Anoto Inc. This pen functions by using a camera to capture an image of paper encoded with a predefined pattern. An example of the image pattern is shown in Figure 11. This pattern is used by the Anoto pen (by Anoto Inc.) to determine a location of a pen on a piece of paper. However, it is unclear how efficient the determination of the location is

with the system used by the Anoto pen. To provide efficient determination of the location of the captured image, a system is needed that provides efficient decoding of the captured image.

- [04] When reviewing a document, a user may annotate the document by moving a pen tip with respect to the document. The path of the pen tip may comprise a plurality of strokes, where each stroke corresponds to a series of captured images. Hence, there is a real need in the industry to identify the path of the pen in order to process the annotation on a document.

Summary

- [05] Aspects of the present invention provide system and methods that determine a path of a pen tip as the pen tip is moved across a document. With an aspect of the invention, the document is watermarked with a maze pattern from which encoded position information is determined.
- [06] With another aspect of the invention, a sequence of images is captured by a camera that is located in a pen. The path of the pen tip is determined by decoding the associated maze pattern (m-array) and by matching the captured images with document images.
- [07] With another aspect of the invention, if the position coordinates of any frame (corresponding to a captured image) that is associated with a stroke cannot be determined from m-array decoding, the frames are transformed and then matched with an area of a document image. Once the position coordinates of at least one frame are determined, the position coordinates of other frames may be determined by matching the frames in a neighboring area.
- [08] With another aspect of the invention, the path of the pen tip (corresponding to a stroke) is mapped from a frame center using a perspective transform and calibration parameters. The perspective transform is obtained from maze pattern analysis and by matching camera-captured images with document images.

Brief Description of Drawings

- [09] The foregoing summary of the invention, as well as the following detailed description of preferred embodiments, is better understood when read in conjunction with the accompanying drawings, which are included by way of example, and not by way of limitation with regard to the claimed invention.
- [10] Figure 1 shows a general description of a computer that may be used in conjunction with embodiments of the present invention.
- [11] Figures 2A and 2B show an image capture system and corresponding captured image in accordance with embodiments of the present invention.
- [12] Figures 3A through 3F show various sequences and folding techniques in accordance with embodiments of the present invention.
- [13] Figures 4A through 4E show various encoding systems in accordance with embodiments of the present invention.
- [14] Figures 5A through 5D show four possible resultant corners associated with the encoding system according to Figures 4A and 4B.
- [15] Figure 6 shows rotation of a captured image portion in accordance with embodiments of the present invention.
- [16] Figure 7 shows various angles of rotation used in conjunction with the coding system of Figures 4A through 4E.
- [17] Figure 8 shows a process for determining the location of a captured array in accordance with embodiments of the present invention.
- [18] Figure 9 shows a method for determining the location of a captured image in accordance with embodiments of the present invention.

- [19] Figure 10 shows another method for determining the location of a captured image in accordance with embodiments of the present invention.
- [20] Figure 11 shows a representation of encoding space in a document according to prior art.
- [21] Figure 12 shows a flow diagram for decoding extracted bits from a captured image in accordance with embodiments of the present invention.
- [22] Figure 13 shows bit selection of extracted bits from a captured image in accordance with embodiments of the present invention.
- [23] Figure 14 shows an apparatus for decoding extracted bits from a captured image in accordance with embodiments of the present invention.
- [24] Figure 15 shows a method for determining a path of a pen-tip from a sequence of captured frames in accordance with an embodiment of the present invention.
- [25] Figure 16 shows an apparatus for determining a path of a pen-tip from a sequence of captured frames in accordance with an embodiment of the present invention.
- [26] Figure 17 shows an example of a captured image in accordance with an embodiment of the invention.
- [27] Figure 18 shows an example of a captured image comprising a text component in accordance with an embodiment of the invention.
- [28] Figure 19 shows a maze pattern analysis of a captured image in accordance with an embodiment of the invention.
- [29] Figure 20 shows a maze pattern analysis of a captured image comprising a text component in accordance with an embodiment of the invention.
- [30] Figure 21 shows a result of analyzing a document image in accordance with an embodiment of the invention.

- [31] Figure 22 shows a result of a global localization for an exemplary stroke in accordance with an embodiment of the invention.
- [32] Figure 23 shows a result of a local localization for an exemplary stroke in accordance with an embodiment of the invention.
- [33] Figure 24 shows a recovered stroke of a pen tip and a path of corresponding centers of captured images in accordance with an embodiment of the invention.
- [34] Figure 25 shows an actual stroke that corresponds to the recovered stroke as shown in Figure 24.

Detailed Description

- [35] Aspects of the present invention relate to determining the location of a captured image in relation to a larger image. The location determination method and system described herein may be used in combination with a multi-function pen.
- [36] The following is separated by subheadings for the benefit of the reader. The subheadings include: terms, general-purpose computer, image capturing pen, encoding of array, decoding, error correction, location determination, m-array decoding, and architecture for determining a path of a pen.

Terms

- [37] Pen - any writing implement that may or may not include the ability to store ink. In some examples, a stylus with no ink capability may be used as a pen in accordance with embodiments of the present invention.
- [38] Camera - an image capture system that may capture an image from paper or any other medium.

General Purpose Computer

- [39] Figure 1 is a functional block diagram of an example of a conventional general-purpose digital computing environment that can be used to implement various aspects of the

present invention. In Figure 1, a computer 100 includes a processing unit 110, a system memory 120, and a system bus 130 that couples various system components including the system memory to the processing unit 110. The system bus 130 may be any of several types of bus structures including a memory bus or memory controller, a peripheral bus, and a local bus using any of a variety of bus architectures. The system memory 120 includes read only memory (ROM) 140 and random access memory (RAM) 150.

[40] A basic input/output system 160 (BIOS), containing the basic routines that help to transfer information between elements within the computer 100, such as during start-up, is stored in the ROM 140. The computer 100 also includes a hard disk drive 170 for reading from and writing to a hard disk (not shown), a magnetic disk drive 180 for reading from or writing to a removable magnetic disk 190, and an optical disk drive 191 for reading from or writing to a removable optical disk 192 such as a CD ROM or other optical media. The hard disk drive 170, magnetic disk drive 180, and optical disk drive 191 are connected to the system bus 130 by a hard disk drive interface 192, a magnetic disk drive interface 193, and an optical disk drive interface 194, respectively. The drives and their associated computer-readable media provide nonvolatile storage of computer readable instructions, data structures, program modules and other data for the personal computer 100. It will be appreciated by those skilled in the art that other types of computer readable media that can store data that is accessible by a computer, such as magnetic cassettes, flash memory cards, digital video disks, Bernoulli cartridges, random access memories (RAMs), read only memories (ROMs), and the like, may also be used in the example operating environment.

[41] A number of program modules can be stored on the hard disk drive 170, magnetic disk 190, optical disk 192, ROM 140 or RAM 150, including an operating system 195, one or more application programs 196, other program modules 197, and program data 198. A user can enter commands and information into the computer 100 through input devices such as a keyboard 101 and pointing device 102. Other input devices (not shown) may include a microphone, joystick, game pad, satellite dish, scanner or the like. These and other input devices are often connected to the processing unit 110 through a serial port

interface 106 that is coupled to the system bus, but may be connected by other interfaces, such as a parallel port, game port or a universal serial bus (USB). Further still, these devices may be coupled directly to the system bus 130 via an appropriate interface (not shown). A monitor 107 or other type of display device is also connected to the system bus 130 via an interface, such as a video adapter 108. In addition to the monitor, personal computers typically include other peripheral output devices (not shown), such as speakers and printers. In a preferred embodiment, a pen digitizer 165 and accompanying pen or stylus 166 are provided in order to digitally capture freehand input. Although a direct connection between the pen digitizer 165 and the serial port is shown, in practice, the pen digitizer 165 may be coupled to the processing unit 110 directly, via a parallel port or other interface and the system bus 130 as known in the art. Furthermore, although the digitizer 165 is shown apart from the monitor 107, it is preferred that the usable input area of the digitizer 165 be co-extensive with the display area of the monitor 107. Further still, the digitizer 165 may be integrated in the monitor 107, or may exist as a separate device overlaying or otherwise appended to the monitor 107.

- [42] The computer 100 can operate in a networked environment using logical connections to one or more remote computers, such as a remote computer 109. The remote computer 109 can be a server, a router, a network PC, a peer device or other common network node, and typically includes many or all of the elements described above relative to the computer 100, although only a memory storage device 111 has been illustrated in Figure 1. The logical connections depicted in Figure 1 include a local area network (LAN) 112 and a wide area network (WAN) 113. Such networking environments are commonplace in offices, enterprise-wide computer networks, intranets and the Internet.
- [43] When used in a LAN networking environment, the computer 100 is connected to the local network 112 through a network interface or adapter 114. When used in a WAN networking environment, the personal computer 100 typically includes a modem 115 or other means for establishing a communications over the wide area network 113, such as the Internet. The modem 115, which may be internal or external, is connected to the system bus 130 via the serial port interface 106. In a networked environment, program

modules depicted relative to the personal computer 100, or portions thereof, may be stored in the remote memory storage device.

- [44] It will be appreciated that the network connections shown are illustrative and other techniques for establishing a communications link between the computers can be used. The existence of any of various well-known protocols such as TCP/IP, Ethernet, FTP, HTTP, Bluetooth, IEEE 802.11x and the like is presumed, and the system can be operated in a client-server configuration to permit a user to retrieve web pages from a web-based server. Any of various conventional web browsers can be used to display and manipulate data on web pages.

Image Capturing Pen

- [45] Aspects of the present invention include placing an encoded data stream in a displayed form that represents the encoded data stream. (For example, as will be discussed with Figure 4B, the encoded data stream is used to create a graphical pattern.) The displayed form may be printed paper (or other physical medium) or may be a display projecting the encoded data stream in conjunction with another image or set of images. For example, the encoded data stream may be represented as a physical graphical image on the paper or a graphical image overlying the displayed image (e.g., representing the text of a document) or may be a physical (non-modifiable) graphical image on a display screen (so any image portion captured by a pen is locatable on the display screen).
- [46] This determination of the location of a captured image may be used to determine the location of a user's interaction with the paper, medium, or display screen. In some aspects of the present invention, the pen may be an ink pen writing on paper. In other aspects, the pen may be a stylus with the user writing on the surface of a computer display. Any interaction may be provided back to the system with knowledge of the encoded image on the document or supporting the document displayed on the computer screen. By repeatedly capturing images with a camera in the pen or stylus as the pen or stylus traverses a document, the system can track movement of the stylus being controlled by the user. The displayed or printed image may be a watermark associated

with the blank or content-rich paper or may be a watermark associated with a displayed image or a fixed coding overlying a screen or built into a screen.

- [47] Figures 2A and 2B show an illustrative example of pen 201 with a camera 203. Pen 201 includes a tip 202 that may or may not include an ink reservoir. Camera 203 captures an image 204 from surface 207. Pen 201 may further include additional sensors and/or processors as represented in broken box 206. These sensors and/or processors 206 may also include the ability to transmit information to another pen 201 and/or a personal computer (for example, via Bluetooth or other wireless protocols).
- [48] Figure 2B represents an image as viewed by camera 203. In one illustrative example, the field of view of camera 203 (i.e., the resolution of the image sensor of the camera) is 32×32 pixels (where $N=32$). In the embodiment, a captured image (32 pixels by 32 pixels) corresponds to an area of approximately 5 mm by 5 mm of the surface plane captured by camera 203. Accordingly, Figure 2B shows a field of view of 32 pixels long by 32 pixels wide. The size of N is adjustable, such that a larger N corresponds to a higher image resolution. Also, while the field of view of the camera 203 is shown as a square for illustrative purposes here, the field of view may include other shapes as is known in the art.
- [49] The images captured by camera 203 may be defined as a sequence of image frames $\{I_i\}$, where I_i is captured by the pen 201 at sampling time t_i . The sampling rate may be large or small, depending on system configuration and performance requirement. The size of the captured image frame may be large or small, depending on system configuration and performance requirement.
- [50] The image captured by camera 203 may be used directly by the processing system or may undergo pre-filtering. This pre-filtering may occur in pen 201 or may occur outside of pen 201 (for example, in a personal computer).
- [51] The image size of Figure 2B is 32×32 pixels. If each encoding unit size is 3×3 pixels, then the number of captured encoded units would be approximately 100 units. If the

encoding unit size is 5x5 pixels, then the number of captured encoded units is approximately 36.

- [52] Figure 2A also shows the image plane 209 on which an image 210 of the pattern from location 204 is formed. Light received from the pattern on the object plane 207 is focused by lens 208. Lens 208 may be a single lens or a multi-part lens system, but is represented here as a single lens for simplicity. Image capturing sensor 211 captures the image 210.
- [53] The image sensor 211 may be large enough to capture the image 210. Alternatively, the image sensor 211 may be large enough to capture an image of the pen tip 202 at location 212. For reference, the image at location 212 is referred to as the virtual pen tip. It is noted that the virtual pen tip location with respect to image sensor 211 is fixed because of the constant relationship between the pen tip, the lens 208, and the image sensor 211.
- [54] The following transformation $F_{S \rightarrow P}$ transforms position coordinates in the image captured by camera to position coordinates in the real image on the paper:

$$L_{paper} = F_{S \rightarrow P}(L_{Sensor})$$

- [55] During writing, the pen tip and the paper are on the same plane. Accordingly, the transformation from the virtual pen tip to the real pen tip is also $F_{S \rightarrow P}$:

$$L_{pentip} = F_{S \rightarrow P}(L_{virtual-pentip})$$

- [56] The transformation $F_{S \rightarrow P}$ may be estimated as an affine transform. This simplifies as:

$$F'_{S \rightarrow P} = \begin{Bmatrix} \frac{s_x \sin \theta_y}{\cos \theta_x \sin \theta_y - \cos \theta_y \sin \theta_x}, & -\frac{s_x \cos \theta_y}{\cos \theta_x \sin \theta_y - \cos \theta_y \sin \theta_x}, & 0 \\ \frac{s_y \sin \theta_x}{\cos \theta_x \sin \theta_y - \cos \theta_y \sin \theta_x}, & \frac{s_y \cos \theta_x}{\cos \theta_x \sin \theta_y - \cos \theta_y \sin \theta_x}, & 0 \\ 0, & 0, & 1 \end{Bmatrix}$$

as the estimation of $F_{S \rightarrow P}$, in which θ_x , θ_y , s_x , and s_y are the rotation and scale of two orientations of the pattern captured at location 204. Further, one can refine $F'_{S \rightarrow P}$ by matching the captured image with the corresponding real image on paper. "Refine" means to get a more precise estimation of the transformation $F_{S \rightarrow P}$ by a type of optimization algorithm referred to as a recursive method. The recursive method treats the matrix $F'_{S \rightarrow P}$ as the initial value. The refined estimation describes the transformation between S and P more precisely.

- [57] Next, one can determine the location of virtual pen tip by calibration.
- [58] One places the pen tip 202 on a fixed location L_{pentip} on paper. Next, one tilts the pen, allowing the camera 203 to capture a series of images with different pen poses. For each image captured, one may obtain the transformation $F_{S \rightarrow P}$. From this transformation, one can obtain the location of the virtual pen tip $L_{virtual-pentip}$:

$$L_{virtual-pentip} = F_{P \rightarrow S}(L_{pentip})$$

where L_{pentip} is initialized as (0, 0) and

$$F_{P \rightarrow S} = (F_{S \rightarrow P})^{-1}$$

- [59] By averaging the $L_{virtual-pentip}$ obtained from each image, a location of the virtual pen tip $L_{virtual-pentip}$ may be determined. With $L_{virtual-pentip}$, one can get a more accurate estimation of L_{pentip} . After several times of iteration, an accurate location of virtual pen tip $L_{virtual-pentip}$ may be determined.
- [60] The location of the virtual pen tip $L_{virtual-pentip}$ is now known. One can also obtain the transformation $F_{S \rightarrow P}$ from the images captured. Finally, one can use this information to determine the location of the real pen tip L_{pentip} :

$$L_{pentip} = F_{S \rightarrow P}(L_{virtual-pentip})$$

Encoding of Array

- [61] A two-dimensional array may be constructed by folding a one-dimensional sequence. Any portion of the two-dimensional array containing a large enough number of bits may be used to determine its location in the complete two-dimensional array. However, it may be necessary to determine the location from a captured image or a few captured images. So as to minimize the possibility of a captured image portion being associated with two or more locations in the two-dimensional array, a non-repeating sequence may be used to create the array. One property of a created sequence is that the sequence does not repeat over a length (or window) n . The following describes the creation of the one-dimensional sequence then the folding of the sequence into an array.

Sequence Construction

- [62] A sequence of numbers may be used as the starting point of the encoding system. For example, a sequence (also referred to as an m -sequence) may be represented as a q -element set in field F_q . Here, $q=p^n$ where $n \geq 1$ and p is a prime number. The sequence or m -sequence may be generated by a variety of different techniques including, but not limited to, polynomial division. Using polynomial division, the sequence may be defined as follows:

$$\frac{R_l(x)}{P_n(x)}$$

- [63] where $P_n(x)$ is a primitive polynomial of degree n in field $F_q[x]$ (having q^n elements). $R_l(x)$ is a nonzero polynomial of degree l (where $l < n$) in field $F_q[x]$. The sequence may be created using an iterative procedure with two steps: first, dividing the two polynomials (resulting in an element of field F_q) and, second, multiplying the remainder by x . The computation stops when the output begins to repeat. This process may be implemented using a linear feedback shift register as set forth in an article by Douglas W. Clark and Lih-Jyh Weng, "Maximal and Near-Maximal Shift Register Sequences: Efficient Event Counters and Easy Discrete Logarithms," IEEE Transactions on Computers 43.5 (May 1994, pp 560-568). In this environment, a relationship is established between cyclical

shifting of the sequence and polynomial $R_l(x)$: changing $R_l(x)$ only cyclically shifts the sequence and every cyclical shifting corresponds to a polynomial $R_l(x)$. One of the properties of the resulting sequence is that, the sequence has a period of $q^n - 1$ and within a period, over a width (or length) n , any portion exists once and only once in the sequence. This is called the “window property”. Period $q^n - 1$ is also referred to as the length of the sequence and n as the order of the sequence.

- [64] The process described above is but one of a variety of processes that may be used to create a sequence with the window property.

Array Construction

- [65] The array (or m -array) that may be used to create the image (of which a portion may be captured by the camera) is an extension of the one-dimensional sequence or m -sequence. Let A be an array of period (m_1, m_2) , namely $A(k + m_1, l) = A(k, l + m_2) = A(k, l)$. When an $n_1 \times n_2$ window shifts through a period of A , all the nonzero $n_1 \times n_2$ matrices over F_q appear once and only once. This property is also referred to as a “window property” in that each window is unique.
- [66] A binary array (or m -array) may be constructed by folding the sequence. One approach is to obtain a sequence then fold it to a size of $m_1 \times m_2$ where the length of the array is $L = m_1 \times m_2 = 2^n - 1$. Alternatively, one may start with a predetermined size of the space that one wants to cover (for example, one sheet of paper, 30 sheets of paper or the size of a computer monitor), determine the area $(m_1 \times m_2)$, then use the size to let $L \geq m_1 \times m_2$, where $L = 2^n - 1$.
- [67] A variety of different folding techniques may be used. For example, Figures 3A through 3C show three different sequences. Each of these may be folded into the array shown as Figure 3D. The three different folding methods are shown as the overlay in Figure 3D and as the raster paths in Figures 3E and 3F. We adopt the folding method shown in Figure 3D.

- [68] To create the folding method as shown in Figure 3D, one creates a sequence $\{a_i\}$ of length L and order n . Next, an array $\{b_{kl}\}$ of size $m_1 \times m_2$, where $\gcd(m_1, m_2) = 1$ and $L = m_1 \times m_2$, is created from the sequence $\{a_i\}$ by letting each bit of the array be calculated as shown by equation 1:

$$b_{kl} = a_i, \text{ where } k = i \bmod(m_1), l = i \bmod(m_2), i = 0, \dots, L-1. \quad (1)$$

- [69] This folding approach may be alternatively expressed as laying the sequence on the diagonal of the array, then continuing from the opposite edge when an edge is reached.
- [70] Figure 4A shows sample encoding techniques that may be used to encode the array of Figure 3D. It is appreciated that other encoding techniques may be used. For example, an alternative coding technique is shown in Figure 11.
- [71] Referring to Figure 4A, a first bit 401 (for example, "1") is represented by a column of dark ink. A second bit 402 (for example, "0") is represented by a row of dark ink. It is appreciated that any color ink may be used to represent the various bits. The only requirement in the color of the ink chosen is that it provides a significant contrast with the background of the medium to be differentiable by an image capture system. The bits in Figure 4A are represented by a 3x3 matrix of cells. The size of the matrix may be modified to be any size as based on the size and resolution of an image capture system. Alternative representation of bits 0 and 1 are shown in Figures 4C-4E. It is appreciated that the representation of a one or a zero for the sample encodings of Figures 4A-4E may be switched without effect. Figure 4C shows bit representations occupying two rows or columns in an interleaved arrangement. Figure 4D shows an alternative arrangement of the pixels in rows and columns in a dashed form. Finally Figure 4E shows pixel representations in columns and rows in an irregular spacing format (e.g., two dark dots followed by a blank dot).
- [72] Referring back to Figure 4A, if a bit is represented by a 3×3 matrix and an imaging system detects a dark row and two white rows in the 3×3 region, then a zero is detected

(or one). If an image is detected with a dark column and two white columns, then a one is detected (or a zero).

- [73] Here, more than one pixel or dot is used to represent a bit. Using a single pixel (or bit) to represent a bit is fragile. Dust, creases in paper, non-planar surfaces, and the like create difficulties in reading single bit representations of data units. However, it is appreciated that different approaches may be used to graphically represent the array on a surface. Some approaches are shown in Figures 4C through 4E. It is appreciated that other approaches may be used as well. One approach is set forth in Figure 11 using only space-shifted dots.
- [74] A bit stream is used to create the graphical pattern 403 of Figure 4B. Graphical pattern 403 includes 12 rows and 18 columns. The rows and columns are formed by a bit stream that is converted into a graphical representation using bit representations 401 and 402. Figure 4B may be viewed as having the following bit representation:

$$\begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

Decoding

- [75] When a person writes with the pen of Figure 2A or moves the pen close to the encoded pattern, the camera captures an image. For example, pen 201 may utilize a pressure sensor as pen 201 is pressed against paper and pen 201 traverses a document on the paper. The image is then processed to determine the orientation of the captured image with respect to the complete representation of the encoded image and extract the bits that make up the captured image.
- [76] For the determination of the orientation of the captured image relative to the whole encoded area, one may notice that not all the four conceivable corners shown in Figure 5A-5D can present in the graphical pattern 403. In fact, with the correct orientation, the

type of corner shown in Figure 5A cannot exist in the graphical pattern 403. Therefore, the orientation in which the type of corner shown in Figure 5A is missing is the right orientation.

- [77] Continuing to Figure 6, the image captured by a camera 601 may be analyzed and its orientation determined so as to be interpretable as to the position actually represented by the image 601. First, image 601 is reviewed to determine the angle θ needed to rotate the image so that the pixels are horizontally and vertically aligned. It is noted that alternative grid alignments are possible including a rotation of the underlying grid to a non-horizontal and vertical arrangement (for example, 45 degrees). Using a non-horizontal and vertical arrangement may provide the probable benefit of eliminating visual distractions from the user, as users may tend to notice horizontal and vertical patterns before others. For purposes of simplicity, the orientation of the grid (horizontal and vertical and any other rotation of the underlying grid) is referred to collectively as the predefined grid orientation.
- [78] Next, image 601 is analyzed to determine which corner is missing. The rotation amount ϕ needed to rotate image 601 to an image ready for decoding 603 is shown as $\phi = (\theta \text{ plus a rotation amount \{defined by which corner missing\}})$. The rotation amount is shown by the equation in Figure 7. Referring back to Figure 6, angle θ is first determined by the layout of the pixels to arrive at a horizontal and vertical (or other predefined grid orientation) arrangement of the pixels and the image is rotated as shown in 602. An analysis is then conducted to determine the missing corner and the image 602 rotated to the image 603 to set up the image for decoding. Here, the image is rotated 90 degrees counterclockwise so that image 603 has the correct orientation and can be used for decoding.
- [79] It is appreciated that the rotation angle θ may be applied before or after rotation of the image 601 to account for the missing corner. It is also appreciated that by considering noise in the captured image, all four types of corners may be present. We may count the

number of corners of each type and choose the type that has the least number as the corner type that is missing.

- [80] Finally, the code in image 603 is read out and correlated with the original bit stream used to create image 403. The correlation may be performed in a number of ways. For example, it may be performed by a recursive approach in which a recovered bit stream is compared against all other bit stream fragments within the original bit stream. Second, a statistical analysis may be performed between the recovered bit stream and the original bit stream, for example, by using a Hamming distance between the two bit streams. It is appreciated that a variety of approaches may be used to determine the location of the recovered bit stream within the original bit stream.
- [81] Once one has the recovered bits, one needs to locate the captured image within the original array (for example, the one shown in Figure 4B). The process of determining the location of a segment of bits within the entire array is complicated by a number of items. First, the actual bits to be captured may be obscured (for example, the camera may capture an image with document content that obscures the code). Second, dust, creases, reflections, and the like may also create errors in the captured image. These errors make the localization process more difficult. In this regard, the image capture system may need to function with non-sequential bits extracted from the image. The following represents a method for operating with non-sequential bits from the image.
- [82] Let the sequence (or m -sequence) \mathbf{I} correspond to the power series $I(x) = 1/P_n(x)$, where n is the order of the m -sequence, and the captured image contains K bits of \mathbf{I} $\mathbf{b} = (b_0 \ b_1 \ b_2 \ \cdots \ b_{K-1})^t$, where $K \geq n$ and the superscript t represents a transpose of the matrix or vector. The location s of the K bits is just the number of cyclic shifts of \mathbf{I} so that b_0 is shifted to the beginning of the sequence. Then this shifted sequence \mathbf{R} corresponds to the power series $x^s/P_n(x)$, or $\mathbf{R} = T^s(\mathbf{I})$, where T is the cyclic shift operator. We find this s indirectly. The polynomials modulo $P_n(x)$ form a field. It is

guaranteed that $x^s \equiv r_0 + r_1x + \cdots r_{n-1}x^{n-1} \pmod{P_n(x)}$. Therefore, we may find $(r_0, r_1, \dots, r_{n-1})$ and then solve for s .

- [83] The relationship $x^s \equiv r_0 + r_1x + \cdots r_{n-1}x^{n-1} \pmod{P_n(x)}$ implies that $\mathbf{R} = r_0 + r_1T(\mathbf{I}) + \cdots + r_{n-1}T^{n-1}(\mathbf{I})$. Written in a binary linear equation, it becomes:

$$\mathbf{R} = \mathbf{r}'\mathbf{A} \quad (2)$$

where $\mathbf{r} = (r_0 \ r_1 \ r_2 \ \cdots \ r_{n-1})'$, and $\mathbf{A} = (\mathbf{I} \ T(\mathbf{I}) \ \cdots \ T^{n-1}(\mathbf{I}))'$ which consists of the cyclic shifts of \mathbf{I} from 0-shift to $(n-1)$ -shift. Now only sparse K bits are available in \mathbf{R} to solve \mathbf{r} . Let the index differences between b_i and b_0 in \mathbf{R} be k_i , $i = 1, 2, \dots, k-1$, then the 1st and $(k_i + 1)$ -th elements of \mathbf{R} , $i = 1, 2, \dots, k-1$, are exactly b_0, b_1, \dots, b_{k-1} . By selecting the 1st and $(k_i + 1)$ -th columns of \mathbf{A} , $i = 1, 2, \dots, k-1$, the following binary linear equation is formed:

$$\mathbf{b}' = \mathbf{r}'\mathbf{M} \quad (3)$$

where \mathbf{M} is an $n \times K$ sub-matrix of \mathbf{A} .

- [84] If \mathbf{b} is error-free, the solution of \mathbf{r} may be expressed as:

$$\mathbf{r}' = \tilde{\mathbf{b}}'\tilde{\mathbf{M}}^{-1} \quad (4)$$

- [85] where $\tilde{\mathbf{M}}$ is any non-degenerate $n \times n$ sub-matrix of \mathbf{M} and $\tilde{\mathbf{b}}$ is the corresponding sub-vector of \mathbf{b} .

- [86] With known \mathbf{r} , we may use the Pohlig-Hellman-Silver algorithm as noted by Douglas W. Clark and Lih-Jyh Weng, "Maximal and Near-Maximal Shift Register Sequences: Efficient Event Counters and Easy Discrete Logarithms," IEEE Transactions on Computers 43.5 (May 1994, pp 560-568) to find s so that $x^s \equiv r_0 + r_1x + \cdots r_{n-1}x^{n-1} \pmod{P_n(x)}$.

- [87] As matrix \mathbf{A} (with the size of n by L , where $L=2^n - 1$) may be huge, we should avoid storing the entire matrix \mathbf{A} . In fact, as we have seen in the above process, given extracted bits with index difference k_i , only the first and $(k_i + 1)$ -th columns of \mathbf{A} are relevant to the computation. Such choices of k_i is quite limited, given the size of the captured image. Thus, only those columns that may be involved in computation need to be saved. The total number of such columns is much smaller than L (where $L=2^n - 1$ is the length of the m -sequence).

Error Correction

- [88] If errors exist in \mathbf{b} , then the solution of \mathbf{r} becomes more complex. Traditional methods of decoding with error correction may not readily apply, because the matrix \mathbf{M} associated with the captured bits may change from one captured image to another.
- [89] We adopt a stochastic approach. Assuming that the number of error bits in \mathbf{b} , n_e , is relatively small compared to K , then the probability of choosing the correct n bits from the K bits of \mathbf{b} and the corresponding sub-matrix $\tilde{\mathbf{M}}$ of \mathbf{M} being non-degenerate is high.
- [90] When the n bits chosen are all correct, the Hamming distance between \mathbf{b}' and $\mathbf{r}'\mathbf{M}$, or the number of error bits associated with \mathbf{r} , should be minimal, where \mathbf{r} is computed via equation (4). Repeating the process for several times, it is likely that the correct \mathbf{r} that results in the minimal error bits can be identified.
- [91] If there is only one \mathbf{r} that is associated with the minimum number of error bits, then it is regarded as the correct solution. Otherwise, if there is more than one \mathbf{r} that is associated with the minimum number of error bits, the probability that n_e exceeds the error correcting ability of the code generated by \mathbf{M} is high and the decoding process fails. The system then may move on to process the next captured image. In another implementation, information about previous locations of the pen can be taken into consideration. That is, for each captured image, a destination area where the pen may be expected next can be identified. For example, if the user has not lifted the pen between two image captures by the camera, the location of the pen as determined by the second image capture should not

be too far away from the first location. Each \mathbf{r} that is associated with the minimum number of error bits can then be checked to see if the location s computed from \mathbf{r} satisfies the local constraint, i.e., whether the location is within the destination area specified.

- [92] If the location s satisfies the local constraint, the X, Y positions of the extracted bits in the array are returned. If not, the decoding process fails.
- [93] Figure 8 depicts a process that may be used to determine a location in a sequence (or m -sequence) of a captured image. First, in step 801, a data stream relating to a captured image is received. In step 802, corresponding columns are extracted from \mathbf{A} and a matrix \mathbf{M} is constructed.
- [94] In step 803, n independent column vectors are randomly selected from the matrix \mathbf{M} and vector \mathbf{r} is determined by solving equation (4). This process is performed Q times (for example, 100 times) in step 804. The determination of the number of loop times is discussed in the section Loop Times Calculation later.
- [95] In step 805, \mathbf{r} is sorted according to its associated number of error bits. The sorting can be done using a variety of sorting algorithms as known in the art. For example, a selection sorting algorithm may be used. The selection sorting algorithm is beneficial when the number Q is not large. However, if Q becomes large, other sorting algorithms (for example, a merge sort) that handle larger numbers of items more efficiently may be used.
- [96] The system then determines in step 806 whether error correction was performed successfully, by checking whether multiple \mathbf{r} 's are associated with the minimum number of error bits. If yes, an error is returned in step 809, indicating the decoding process failed. If not, the position s of the extracted bits in the sequence (or m -sequence) is calculated in step 807, for example, by using the Pohig-Hellman-Silver algorithm.
- [97] Next, the (X,Y) position in the array is calculated as: $x = s \bmod m_1$ and $y = s \bmod m_2$ and the results are returned in step 808.

Location Determination

- [98] Figure 9 shows a process for determining the location of a pen tip. The input is an image captured by a camera and the output may be a position coordinates of the pen tip. Also, the output may include (or not) other information such as a rotation angle of the captured image.
- [99] In step 901, an image is received from a camera. Next, the received image may be optionally preprocessed in step 902 (as shown by the broken outline of step 902) to adjust the contrast between the light and dark pixels and the like.
- [100] Next, in step 903, the image is analyzed to determine the bit stream within it.
- [101] Next, in step 904, n bits are randomly selected from the bit stream for multiple times and the location of the received bit stream within the original sequence (or m -sequence) is determined.
- [102] Finally, once the location of the captured image is determined in step 904, the location of the pen tip may be determined in step 905.
- [103] Figure 10 gives more details about 903 and 904 and shows the approach to extract the bit stream within a captured image. First, an image is received from the camera in step 1001. The image then may optionally undergo image preprocessing in step 1002 (as shown by the broken outline of step 1002). The pattern is extracted in step 1003. Here, pixels on the various lines may be identified and the orientation of the pattern (corresponding to the angle θ) passing through the pixel can be estimated.
- [104] Next, the received image is analyzed in step 1004 to determine the underlying grid lines. If grid lines are found in step 1005, then the code is extracted from the pattern in step 1006. The code is then decoded in step 1007 and the location of the pen tip is determined in step 1008. If no grid lines were found in step 1005, then an error is returned in step 1009.

Outline of Enhanced Decoding and Error Correction Algorithm

[105] With an embodiment of the invention as shown in Figure 12, given extracted bits 1201 from a captured image (corresponding to a captured array) and the destination area, a variation of an m -array decoding and error correction process decodes the X,Y position. Figure 12 shows a flow diagram of process 1200 of this enhanced approach. Process 1200 comprises two components 1251 and 1253.

- ***Decode Once.*** Component 1251 includes three parts.
 - random bit selection: randomly selects a subset of the extracted bits 1201 (step 1203)
 - decode the subset (step 1205)
 - determine X,Y position with local constraint (step 1209)
- ***Decoding with Smart Bit Selection.*** Component 1253 includes four parts.
 - smart bit selection: selects another subset of the extracted bits (step 1217)
 - decode the subset (step 1219)
 - adjust the number of iterations (loop times) of step 1217 and step 1219 (step 1221)
 - determine X,Y position with local constraint (step 1225)

[106] The embodiment of the invention utilizes a discreet strategy to select bits, adjusts the number of loop iterations, and determines the X,Y position (location coordinates) in accordance with a local constraint, which is provided to process 1200. With both components 1251 and 1253, steps 1205 and 1219 (“Decode Once”) utilize equation (4) to compute \mathbf{r} .

Let $\hat{\mathbf{b}}$ be decoded bits, that is:

$$\hat{\mathbf{b}}' = \mathbf{r}'\mathbf{M} \quad (5)$$

The difference between \mathbf{b} and $\hat{\mathbf{b}}$ are the error bits associated with \mathbf{r} .

[107] Figure 12 shows a flow diagram of process 1200 for decoding extracted bits 1201 from a captured image in accordance with embodiments of the present invention. Process 1200 comprises components 1251 and 1253. Component 1251 obtains extracted bits 1201 (comprising K bits) associated with a captured image (corresponding to a captured array). In step 1203, n bits (where n is the order of the m -array) are randomly selected from extracted bits 1201. In step 1205, process 1200 decodes once and calculates r . In step 1207, process 1200 determines if error bits are detected for \mathbf{b} . If step 1207 determines that there are no error bits, X,Y coordinates of the position of the captured array are determined in step 1209. With step 1211, if the X,Y coordinates satisfy the local constraint, i.e., coordinates that are within the destination area, process 1200 provides the X,Y position (such as to another process or user interface) in step 1213. Otherwise, step 1215 provides a failure indication.

[108] If step 1207 detects error bits in \mathbf{b} , component 1253 is executed in order to decode with error bits. Step 1217 selects another set of n bits (which differ by at least one bit from the n bits selected in step 1203) from extracted bits 1201. Steps 1221 and 1223 determine the number of iterations (loop times) that are necessary for decoding the extracted bits. Step 1225 determines the position of the captured array by testing which candidates obtained in step 1219 satisfy the local constraint. Steps 1217-1225 will be discussed in more details.

Smart Bit Selection

[109] Step 1203 randomly selects n bits from extracted bits 1201 (having K bits), and solves for \mathbf{r}_1 . Using equation (5), decoded bits can be calculated. Let $I_1 = \{k \in \{1, 2, \dots, K\} \mid b_k = \hat{b}_k\}$, $\bar{I}_1 = \{k \in \{1, 2, \dots, K\} \mid b_k \neq \hat{b}_k\}$, where \hat{b}_k is the k^{th} bit of $\hat{\mathbf{b}}$, $B_1 = \{b_k \mid k \in I_1\}$ and $\bar{B}_1 = \{b_k \mid k \in \bar{I}_1\}$, that is, B_1 are bits that the decoded results are the same as the original bits, and \bar{B}_1 are bits that the decoded results are different from the original bits, I_1 and \bar{I}_1 are the corresponding indices of these bits. It is appreciated that the same \mathbf{r}_1 will be obtained when any n bits are selected from B_1 . Therefore, if the next n bits are not

carefully chosen, it is possible that the selected bits are a subset of B_1 , thus resulting in the same r_1 being obtained.

[110] In order to avoid such a situation, step 1217 selects the next n bits according to the following procedure:

1. Choose at least one bit from \bar{B}_1 1303 and the rest of the bits randomly from B_1 1301 and \bar{B}_1 1303, as shown in Figure 13 corresponding to bit arrangement 1351. Process 1200 then solves r_2 and finds B_2 1305, 1309 and \bar{B}_2 1307, 1311 by computing $\hat{b}_2^t = r_2^t M_2$.
2. Repeat step 1. When selecting the next n bits, for every \bar{B}_i ($i = 1, 2, 3, \dots, x-1$, where x is the current loop number), there is at least one bit selected from \bar{B}_i . The iteration terminates when no such subset of bits can be selected or when the loop times are reached.

Loop Times Calculation

[111] With the error correction component 1253, the number of required iterations (loop times) is adjusted after each loop. The loop times is determined by the expected error rate. The expected error rate p_e in which not all the selected n bits are correct is:

$$p_e = \left(1 - \frac{C_{K-n_e}^n}{C_K^n} \right)^{lt} \approx -e^{-lt \left(\frac{K-n}{K} \right)^n} \quad (6)$$

where lt represents the loop times and is initialized by a constant, K is the number of extracted bits from the captured array, n_e represents the minimum number of error bits incurred during the iteration of process 1200, n is the order of the m -array, and C_K^n is the number of combinations in which n bits are selected from K bits.

[112] In the embodiment, we want p_e to be less than $e^{-5} = 0.0067$. In combination with (6), we have:

$$lt_i = \min \left(lt_{i-1}, \frac{5}{\left(\frac{K-n}{K} \right)^{n_i}} + 1 \right) \quad (7)$$

Adjusting the loop times may significantly reduce the number of iterations of process 1253 that are required for error correction.

Determine X, Y Position with Local Constraint

- [113] In steps 1209 and 1225, the decoded position should be within the destination area. The destination area is an input to the algorithm, and it may be of various sizes and places or simply the whole m -array depending on different applications. Usually it can be predicted by the application. For example, if the previous position is determined, considering the writing speed, the destination area of the current pen tip should be close to the previous position. However, if the pen is lifted, then its next position can be anywhere. Therefore, in this case, the destination area should be the whole m -array. The correct X,Y position is determined by the following steps.

- [114] In step 1224 process 1200 selects r_i whose corresponding number of error bits is less than:

$$N_e = \frac{\log_{10} \left(\frac{3}{lt} \right)}{\log_{10} \left(\frac{K-n}{K} \right) \times \log_{10} \left(\frac{10}{lr} \right)} \quad (8)$$

where lt is the actual loop times and lr represents the Local Constraint Rate calculated by:

$$lr = \frac{\text{area of the destination area}}{L} \quad (9)$$

where L is the length of the m -array.

- [115] Step 1224 sorts r_i in ascending order of the number of error bits. Steps 1225, 1211 and 1212 then finds the first r_i in which the corresponding X,Y position is within the destination area. Steps 1225, 1211 and 1212 finally returns the X,Y position as the result (through step 1213), or an indication that the decoding procedure failed (through step 1215).

Architecture for Determining Path of a Stroke (Strokes Localization)

- [116] Figure 15 shows a method for determining a path of a pen-tip from a sequence of captured frames 1551 in accordance with an embodiment of the present invention. In step 1501, a frame is processed so that document contents such as text are separated from other areas that contain only maze pattern cells. Also, the image (gray-scale) is normalized to compensate for non-uniform illumination. In step 1503, m-array bits are extracted from visible maze pattern bars in a frame (captured image). In step 1505, if the number of extracted bits (as determined in step 1503) is greater than the order of the embedded m-array, unique position coordinates (corresponding to the x-y position of the center of the frame) may be obtained by m-array decoding.
- [117] Step 1507 analyzes digital document 1553 in order to determine the areas of document 1553 in which the maze pattern is occluded by contents of document 1553. (With the embodiment, digital document 1553 does not include annotations generated by a user.) If maze pattern cells are occluded by the contents of document 1553, step 1505 may not be able to extract enough m-array bits to determine the x-y position of a frame. By analyzing the whole watermarked document 1553, process 1500 can ascertain the areas of document 1553 where the x-y position cannot be uniquely determined. This analysis may be used in step 1509 as will be discussed.
- [118] In step 1511, if the x-y position of any frame that is associated with a stroke cannot be determined from m-array decoding (corresponding to step 1505), process 1500 may not be able to locate the position of the frame. In such cases, step 1509 performs global localization in which captured images can be warped by an affine transform obtained by maze pattern analysis (step 1503) and then can be matched with an area of document image 1553 where a substantial amount of maze pattern cells are occluded by the content of the document. Once one frame is successfully matched, local localization (step 1517 as will be discussed) is utilized to locate the entire stroke (corresponding to a series of frames).

- [119] In step 1517, frames whose x-y positions are not decoded in step 1505 or determined by global localization in step 1509 are further processed. The location of such frames should be near the location of frames whose x-y positions are determined by m-array decoding or global localization. The location of such frames is determined by matching the frames with the neighboring area of already determined x-y positions. The perspective transform $F_{S \rightarrow P}$ (as previously discussed) between captured images and document images is also obtained.
- [120] X-y positions obtained from m-array decoding (corresponding to step 1505) and global/local localization (steps 1509 and 1517, respectively) represent the locations (x-y positions) of the centers of captured images. (An embodiment of the invention may support one, two or three dimensions. In the embodiment, two dimensions are supported so that a position location corresponds to an x-y position.) In order to obtain the x-y positions of the pen tip, a relationship between the pen tip (e.g., pen tip 202 as shown in Figure 2) and the associated camera (e.g., camera 203 as shown in Figure 2) may need calibration. Step 1513 supports calibration as will be discussed.
- [121] In step 1515 the x-y positions of the pen tip are determined by mapping the x-y positions of image centers using the perspective transform obtained from local localization and calibration parameters. (An example of determining a path 1555 of a pen tip is shown in Figure 26 as will be discussed.)
- [122] Figure 16 shows an apparatus 1600 for determining a path of a pen tip from a sequence of captured images (frames) in accordance with an embodiment of the present invention. In an embodiment of the invention, image segmentation 1601 performs step 1501, maze pattern analyzer 1603 performs step 1503, m-array decoding module 1605 performs step 1505, matching module 1609 performs steps 1511, 1509, and 1517, watermarked document analyzer 1607 performs step 1507, calibration module 1611 performs step 1513, and mapping module 1613 performs step 1515.
- [123] Document image analysis and calibration may be performed off-line, or separately before on-line image-capturing and processing. The other components (e.g., image segmentation

module 1601, maze pattern analyzer 1603, m-array decoding module 1605, matching module 1609, and mapping module 1613) may be performed on-line or off-line. The functionality of the components is described in detail as follows.

Image segmentation

- [124] Captured images may contain contents of a document such as text or drawings, which are initially assessed in preprocessing. Figure 17 shows an example of a captured image 1700 in accordance with an embodiment of the invention. Figure 18 shows an example of a captured image comprising a text component (that includes text characters 1801 and 1803) in accordance with an embodiment of the invention. If an image contains a text/drawing area, the text/drawing area is separated from the other area that contains only maze patterns or blank pixels.
- [125] In addition, the captured images (e.g., captured images 1700 and 1800) are normalized to compensate for the non-uniform illumination. Figure 19 shows a processed image 1900 corresponding to a normalization of illumination of captured image 1700 (as shown in Figure 17) in accordance with an embodiment of the invention. Figure 20 shows a processed image 2000 corresponding to a normalization of illumination of captured image 1800 in accordance with an embodiment of the invention.

Maze pattern analysis

- [126] The task of maze pattern analysis is to extract m-array bits from the visible maze pattern bars (e.g., bars 401 and 402 as shown in Figure 4A) in a captured image. Figure 19 shows a maze pattern analysis 1900 of captured image 1700 (as shown in Figure 17) in accordance with an embodiment of the invention. Bit 1901 (corresponding to a bit value 1902 of "0") and bit 1903 (corresponding to a bit value of 1904 of "1") are two members of the m-array bits. The m-array bits are organized in the maze pattern according to maze pattern grid lines, e.g., grid lines 1905 and 1907. Figure 20 shows a maze pattern analysis 2000 of captured image 1800 (as shown in Figure 18) in accordance with an embodiment of the invention. (Note that m-array bits of some of the maze pattern cells in the vicinity of text characters 1801 and 1803 may not be determined in the example.)

- [127] Figures 19 and 20 show an illustration of maze pattern cells and maze pattern bars. The parameters of maze pattern grid lines (scale and rotation along each dimension, i.e. affine transform) are first calculated, and then the original direction (or quadrant) in which the maze pattern is embedded is determined. Consequently, the m-array bits information is determined based on the grid lines and bar directions.

m-array decoding

- [128] If the number of extracted bits obtained in step 1505 (as shown in Figure 15) is greater than the order of embedded m-array, a unique x-y position may be obtained by m-array decoding.

Watermarked document image analysis

- [129] Maze patterns are possibly occluded by content of the document, which means there may not be enough m-array bits that can be extracted for decoding from a captured image. By analyzing the whole watermarked document image, process 1500 or apparatus 1600 can determine in which area where the x-y position can be uniquely determined and in which area where x-y position cannot be uniquely determined. The analysis result is used in the global localization. Figure 21 shows a result 2100 of analyzing a document image 1553 in accordance with an embodiment of the invention. In the embodiment, pixels on document images (e.g., document 1553) are labeled as one of four types. Pixels are labeled according to the following association:

Type I: 32 by 32 pixel neighboring window (with the pixel under consideration as the center) contains only maze pattern cells

Type II: 32 by 32 pixel neighboring window contains 60 or more maze pattern cells

Type III: 32 by 32 pixel neighboring window contains 36 to 60 maze pattern cells

Type IV: 32 by 32 pixel neighboring window contains 35 or less maze pattern cells

In the embodiment, the x-y position of a captured image can be determined if the center of a captured image is located in type I or type II areas and may be determined if the

center is located in a type III area. In Figure 21, area 2101 corresponds to a type I pixel area, area 2103 corresponds to a type II pixel area, area 2107 corresponds to a type III pixel area, and area 2105 corresponds to a type IV pixel area. (The entire document is analyzed and labeled but only a portion is shown in Figure 21.)

- [130] In the embodiment, process 1500 may obtain digital document 1553 by rendering an electronic document to a bitmap or by scanning a paper document and then re-sampling the corresponding bitmap to an appropriate resolution. Determination of the resolution is based on the following considerations: (1) the resolution of the document image shouldn't be less than that of the captured image, because the watermarked gray-scale document image will be matched with a camera-captured image in order to determine the location of the captured image; (2) one printed maze pattern cell should be mapped to an integer number of document image pixels so that the matching algorithm can work more effectively. For example, if the resolution of the camera is 0.15mm/pixel, i.e., 0.15mm in the physical world is mapped to one camera pixel, the printed maze pattern cell size is 0.45mm*0.45mm, i.e., a printed maze pattern cell is mapped to 3*3 pixels on the camera sensor, the resolution of the document image should also be set to 0.15mm/pixel so that a printed maze pattern cell will be mapped to a 3*3 pixel area in the document image.

Global localization by fast image match

- [131] If the x-y positions of the captured images of a stroke fail to be determined from m-array decoding, it is predictable that the centers of all images are located at the area where x-y position cannot be uniquely determined. In this case, the captured images are warped by affine transform obtained by maze pattern analysis, and then are matched with the area of document image where x-y position cannot be uniquely determined. Once one frame is matched successfully, the local localization algorithm is used to locate the whole stroke.
- [132] With document 1553 (as shown in Figure 15), the number of visible maze pattern cells in a 32 by 32 pixel sub-window typically varies from 10 to 100. Pixels in the document are labeled with one of four types by watermarked document analysis (step 1507 as shown in Figure 15). A search region is set as the collection of type III and type IV areas.

- [133] Figure 22 shows a result of a global localization for an exemplary stroke 2201 in accordance with an embodiment of the invention. The x-y position of the point 2203 on stroke 2201 is determined by global localization (corresponding to step 1509). Portion 2202 of a document is magnified to magnified portion 2204 to better show point 2203 on stroke. (In the embodiment, only one point on stroke 2201 is determined by global localization, and the remaining points are determined by local localization.)
- [134] With global localization, a frame is matched with the document image at select points of the search region. The original frame may be warped (transformed) by the scale and rotation (affine transform) from maze pattern analysis (corresponding to step 1503 as shown in Figure 15). An offset may be useful for determining select points in matching a frame.
- [135] In the embodiment, the success of matching a frame corresponds to the largest cross-correlation value between the captured image and the document image. (The cross-correlation value varies between 0 to 1.) The largest cross-correlation value should be sufficiently larger than other cross-correlation values and the threshold, i.e., how much larger, may be determined by off-line training. For example, global localization may require that the difference between the highest cross-correlation value and the second highest cross-correlation value be larger than 0.1.

Local localization by fast image match

- [136] For those frames which x-y positions are not decoded/determined by m-array decoding/global localization, the locations should be near the locations of the frames where x-y positions are determined. Consequently, the corresponding locations are determined by matching the images with the neighbor areas of already determined x-y positions. The perspective transform between captured images and document images is also obtained.
- [137] Figure 23 shows a result of a local localization for an exemplary stroke 2301 in accordance with an embodiment of the invention. The location positions of a subset of points (e.g., points 2305a and 2305b) on stroke 2301 are determined by m-array decoding

and the positions of the remaining points (e.g., points 2303a and 2303b) are determined by local localization.

[138] In the embodiment, local localization utilizes start points and pivotal start points. Start points are the location of frames which are successfully located by m-array decoding (corresponding to step 1505 as shown in Figure 15) or by global localization (corresponding to step 1509 as shown in Figure 15). A frame sequence is segmented into segments in which a local search is performed segment by segment. A segment may be split into several segments during a local localization. In the embodiment, a local search is restricted to be within a search region that is based upon a pen tip's motion restrictions of velocity and acceleration.

[139] The following procedure depicts an embodiment for local localization:

- **Frame sequence segmentation**

- (a) Select pivotal start points

The first and the last start point in a stroke are pivotal start points.

For the start points between two pivotal start points p_1 , p_2 , the one with the maximum distance D_{\max} to the line L that passes through p_1 , p_2 , and D_{\max} that is larger than a threshold (typically set at 0.5), is a pivotal start point.

- (b) Refine the perspective transform at pivotal start points

Get more accurate perspective transform by comparing the camera-captured image with the document image.

- (c) Segment the frame sequence by pivotal start points

Every segment starts from a pivotal start point or the first point of the stroke and ends at a pivotal start point or the last point of the stroke.

- **Determine finished segment**

A segment is declared a finished segment when:

- (a) There is at least one start point in this segment which is not a pivotal start point, i.e., no more pivotal start points can be found for this segment or in other words, the segment

is very much like a straight line. In this case, all the points are interpolated and the segment is declared finished.

or

(b) For every point in the segment, the corresponding frame has been processed.

- **Find a search region for a frame of unfinished segment**

(a) The first point of the segment is a start point, and need not be processed.

(b) For the second point of the segment, the center of the search region is set to be the first point, and the size of search region is restricted by maximal velocity.

(c) For other points of the segment, the velocity at previous point that has been processed can be estimated. The center of the search region can be calculated from the location and velocity of the previous point, and the size of search region is restricted by maximal acceleration.

- **Template match in search region**

This step is based on the assumption that there is only a little change of a pen's pose in a short period of time, which means that there is only a little difference of the perspective transform between adjacent frames.

(a) Warp the frame by the perspective matrix of previous processed frame.

(b) Determine the cross-correlation at every point of the search region by matching the frame with the document image with the point as the center.

(c) The point with maximum cross-correlation should be the correct location of this frame.

- **Refine perspective transform for the point**

Refine the perspective transform by comparing the camera captured image with the document image.

- **Prune the search result**

Several factors may cause error results to be obtained, such as incorrect initial perspective transform parameters and motion blur. By the motion restriction of velocity and acceleration, one can prune the error results.

(a) All the start points should not be pruned.

- (b) Go through the whole stroke from the first point to the last point. If a point cannot satisfy the motion restriction with the previous points, then this point must be wrong and should be pruned.
- (c) Go through the whole stroke from the last point to the first point. If a point cannot satisfy the motion restriction with the previous points, then this point must be wrong and should be pruned.
- (d) After the pruning, points kept in the stroke are correct points. The pruned points are replaced by points determined from interpolation.

Camera-pen-tip calibration

- [140] X-y positions obtained from m-array decoding and global/local localization represents the locations of the centers of captured images. To obtain the x-y positions of pen tip 202, the relationship between pen tip 202 and the center of camera 203 needs to be calibrated. Fast and accurate calibration is significant since ink cartridges may be changed frequently.
- [141] By touching the pen-tip at a fixed contact point on the writing surface in different postures, a few images are captured. The calibration parameters are determined by using the constraint that all x-y positions of these images should be mapped to the same point (the fixed contact point) by correct calibration parameters.
- [142] In the embodiment, the following procedure is used for calibrating a pen tip. The procedure is used to estimate the calibration parameter $L_{virtual-penip}$:

- a) Place the real pen tip at a fixed location L_{penip} on paper
- b) Keep the pen tip at the location L_{penip} standing, and capture a series of images with different pen poses.
- c) For each captured image, the transform $F_{S \rightarrow P}$ (which transforms the position coordinates in the image captured by camera to position coordinates in the real image on the paper) and $F_{P \rightarrow S} = F_{S \rightarrow P}^{-1}$ (the reverse of $F_{S \rightarrow P}$) is calculated by maze pattern analysis, m-array decoding and matching the captured image with the document image, then:

$$L_{penip}^i = F_{S \rightarrow P}^i \cdot L_{virtual-penip}, i = 1, 2, \dots, N$$

$$L_{virtual-pentip}^i = F_{P \rightarrow S}^i \cdot (L_{pentip} + \Delta L_i), i = 1, 2, \dots, N$$

where N is the number of captured images in the experiment, and ΔP_i is the offset between the actual pen tip location in i^{th} frame and L_{pentip} .

- d) Initialize $\bar{L}_{virtual-pentip}$ as (0, 0), where $\bar{L}_{virtual-pentip}$ is the estimated value of $L_{virtual-pentip}$.
- e) Using the first equation in (c), set $L_{virtual-pentip}$ as $\bar{L}_{virtual-pentip}$ to get L_{pentip}^i , $i = 1, 2, \dots, N$. By averaging L_{pentip}^i , L_{pentip} is estimated as:

$$\bar{L}_{pentip} = \frac{1}{N} \sum_{i=1}^{i=N} L_{pentip}^i$$

- f) Using the second equation in (c), set L_{pentip} as \bar{L}_{pentip} to get $L_{virtual-pentip}^i$, $i = 1, 2, \dots, N$. By averaging $L_{virtual-pentip}^i$, $L_{virtual-pentip}$ is estimated as:

$$\bar{L}_{virtual-pentip} = \frac{1}{N} \sum_{i=1}^{i=N} L_{virtual-pentip}^i$$

- g) Repeat step e. After several iterations, $L_{virtual-pentip}$ and L_{pentip} will converge respectively to substantially more accurate results, which may be referred to as $L_{virtual-pentip}^{accurate}$ and $L_{pentip}^{accurate}$.

Finally, one obtains $L_{virtual-pentip}^{accurate}$ as the calibration parameter $L_{virtual-pentip}$.

Map images' center position to pen tip

- [143] X-y positions of pen tip 202 are determined by mapping the x-y positions of image centers using the perspective transform obtained from local localization and calibration parameters. Figures 24 and 25 illustrate the result of pen tip calibration of an example. Figure 26 shows a recovered stroke 2401 of pen tip 202 in accordance with an embodiment of the invention. Figure 25 shows an actual stroke 2501 that is associated with the recovered stroke as shown in Figure 24. Stroke 2403 in Figure 24 shows the path of the locations of the center of the captured images. That is, without calibration, stroke 2501 may be wrongly recovered (recovered as stroke 2403), whereas with proper calibration, stroke 2501 is correctly recovered (recovered as stroke 2401).
- [144] As can be appreciated by one skilled in the art, a computer system with an associated computer-readable medium containing instructions for controlling the computer system can be utilized to implement the exemplary embodiments that are disclosed herein. The

computer system may include at least one computer such as a microprocessor, digital signal processor, and associated peripheral electronic circuitry.

- [145] Although the invention has been defined using the appended claims, these claims are illustrative in that the invention is intended to include the elements and steps described herein in any combination or sub combination. Accordingly, there are any number of alternative combinations for defining the invention, which incorporate one or more elements from the specification, including the description, claims, and drawings, in various combinations or sub combinations. It will be apparent to those skilled in the relevant technology, in light of the present specification, that alternate combinations of aspects of the invention, either alone or in combination with one or more elements or steps defined herein, may be utilized as modifications or alterations of the invention or as part of the invention. It may be intended that the written description of the invention contained herein covers all such modifications and alterations.